Anexo I

D José Manuel Vergara Martín SECRETARIO DEL DEPARTAMENTO DE BIOLOGÌA DE LA UNIVERSIDAD DE LAS PALMAS DE GRAN CANARIA,

CERTIFICA,

Que el Consejo de Doctores del Departamento en su sesión de fecha.....tomó el acuerdo de dar el consentimiento para su tramitación, a la tesis doctoral titulada “Abundance and diversity in marine microbiol eukaryotes” presentada por el doctorando D Massimo Ciro Pernice y dirigida por el Doctor Ramon Massana i Molera

Y para que así conste, y a efectos de lo previsto en el Artº 6 del Reglamento para la elaboración, defensa, tribunal y evaluación de tesis doctorales de la Universidad de Las Palmas de Gran Canaria, firmo la presente en Las Palmas de Gran Canaria, a...de.....de dos mil.....

Abundance and diversity of marine microbial eukaryotes
(Abundacia y diversidad de eucariotas microbianos marinos)

Massimo Ciro Pernice

Tesis Doctoral presentada por Massimo Ciro Pernice para obtener el grado de Doctor por la Universidad de las Palmas de Gran Canaria, Departamento de Biología, Programa en Oceanografía
(Biennio 2008-2010)

Director: Ramon Massana i Molera

Universidad de las Palmas de Gran Canaria
Institut de Ciències del Mar (ICM-CSIC)

El Doctorando
Massimo Ciro Pernice

El director
Ramon Massana i Molera

En Barcelona, a de de 2014

A mi Familia filogénetica y
a mi Comunidad biogeográfica

Este libro en su conjunto no es más que un borrador; mejor dicho, el borrador de un borrador.
¡Oh, tiempo, fuerza, dinero y paciencia!

(Herman Melville, Moby Dick)

Contents

Summary/Resumen/Resum 13

Introduction 19

Aims and outline 31

Chapter 1 Sequence diversity and novelty of natural assemblages of picoeukaryotes from the Indian Ocean 37

Chapter 2 General patterns of diversity in major marine microeukaryote lineages 63

Chapter 3 Global abundance of planktonic heterotrophic protists in the deep ocean 95

Chapter 4 Diversity of marine microeukaryotes in the global deep ocean 121

Synthesis of results and general discussion 151

Resumen en español 163

General references 213

Agradecimientos 223

Summary

Microeukaryotes are important ecological players in any kind of ecosystem, most notably in the ocean, and it is therefore essential to collect information about their abundance and diversity. To achieve this general goal this thesis was structured in two parts. The first part represents an effort to define our "diversity unit" from studies based on the well-known cloning and Sanger sequencing approach. Basically, we wanted to establish a solid baseline for the second part of the thesis. We started with data from one cruise (Chapter 1) and then continued with the analysis of the complete dataset of 18S rDNA sequences available at that time (Chapter 2). From this analysis we found that the V4 region of the 18S rDNA was a good proxy of the variability of the entire gene. We also determined that the maximal genetic distance for sequences belonging to a same class was 0.25. Once defined this framework, it was used in the second part of the thesis for studying deep ocean microeukaryotes. Thanks to the Malaspina 2010 expedition, we had a comprehensive set of deep samples with associated abiotic and biotic parameters from all over the world. We found that the microeukaryotes abundance averaged 54 cells mL^{-1} in the mesopelagic layer and 14 cells mL^{-1} in the bathypelagic layer, and its variability was explained by depth, prokaryotes abundance and oxygen concentration (Chapter 3). Finally, the diversity of deep microeukaryotes was determined by pyrosequencing and metagenomic tags (Chapter 4). The bathypelagic community was mainly composed by Collodaria, Chrysophyceae, MALV-II and Basidiomycota. However, the relative abundance of these classes varies a lot among samples. The variability in community composition between samples was well explained by the water mass they belong and by the abundance ratio between prokaryotes and microeukaryotes.

Resumen

Los Microeucariotas son actores ecológicos importantes en cualquier tipo de ecosistema, sobre todo en el océano, por lo que es esencial recopilar información acerca de su abundancia y diversidad. Para lograr este objetivo general esta tesis se ha estructurado en dos partes. La primera parte representa un esfuerzo para definir nuestra "unidad de diversidad", empezando por estudios basados en la clonación molecular y la secuenciación de Sanger. Básicamente, queríamos establecer una base sólida para la segunda parte de la tesis. Empezamos con los datos de una campaña (Capítulo 1) y luego seguimos con el análisis del conjunto completo de datos de secuencias de 18S ADNr disponibles en ese momento (Capítulo 2). A partir de este análisis, se encontró que la región V4 del 18S ADNr es un buen indicador de la variabilidad de todo el gen. También se determinó que la distancia genética máxima para las secuencias que pertenecen a una misma clase es de 0.25. Una vez definido este marco, fue utilizado en la segunda parte de la tesis para estudiar los microeucariotas del oceáno profundo. Gracias a la expedición Malaspina 2010, disponíamos de un amplio conjunto de muestras de profundidad de todo el mundo y de sus parámetros abióticos y bióticos asociados. Se encontró que la abundancia de microeucariotas promedio era de 54 células mL^{-1} en la capa mesopelágica y de 14 células ml^{-1} en la capa batipelágica. Su variabilidad se explicaba por la profundidad, la abundancia de procariotas y la concentración de oxígeno (Capítulo 3). Por último, la diversidad de microeucariotas profundos se determinó mediante pirosecuenciación y secuencias de metagenómica (Capítulo 4). La comunidad batipelágica estaba compuesta principalmente por Collodaria, Chrysophyceae, MALV-II y Basidiomycota. Sin embargo, la abundancia relativa de estas clases varía mucho entre las muestras. La variabilidad en la composición de la comunidad entre las muestras se explicaba bien teniendo en cuenta la masa de agua a la que pertenecían y el ratio de abundancia entre procariotas y microeucariotas .

Resum

Els microeucariotes són importants actors ecològics en qualsevol tipus d'ecosistema, sobretot en l'oceà, per la qual cosa és essencial recopilar informació sobre la seva abundància i diversitat. Per aconseguir aquest objectiu general aquesta tesi s'ha estructurat en dues parts. La primera part representa un esforç per definir la nostra "unitat de diversitat" començant per estudis basats en la clonació molecular i seqüenciació de Sanger. Bàsicament, volíem establir una base sòlida per a la segona part de la tesi. Hem començat amb les dades d'un creuer (Capítol 1) i després hem seguit amb l'anàlisi del conjunt complet de dades de seqüències de 18S ADNr disponibles en aquest moment (Capítol 2). A partir d'aquesta anàlisi, es va trobar que la regió V4 del 18S ADNr és un bon indicador de la variabilitat de tot el gen. També es va determinar que la distància genètica màxima per a les seqüències que pertanyen a una mateixa classe és de 0.25. Un cop definit aquest marc, va ser utilitzat en la segona part de la tesi per estudiar microeucariotes de l'oceà profund. Gràcies a l'expedició Malaspina 2010, teníem un ampli conjunt de mostres profundes associades a paràmetres abiòtics i biòtics d'arreu del món. Es va trobar que l'abundància mitjana de microeucariotes era de 54 cèl·lules mL^{-1} a la capa mesopelàgica i 14 cèl·lules mL^{-1} a la capa batipelàgica, i la seva variabilitat s'explicava per la profunditat, l'abundància de procariotes i la concentració d'oxigen (Capítol 3). Finalment, la diversitat de microeucariotes de l'oceà profund es va determinar mitjançant piroseqüenciació i seqüències de metagenòmica (Capítol 4). La comunitat batipelàgica estava composta principalment per Collodaria, Chrysophyceae, MALV-II i Basidiomycota. No obstant això, l'abundància relativa d'aquestes classes varia molt entre les mostres. La variabilitat en la composició de la comunitat entre les mostres queda ben explicada per la massa d'aigua de pertinença i per la ràtio d'abundància entre procariotes i microeucariotes.

Introduction

Marine protist research: a brief history

There is an extremely broad diversity of organisms that fall within the term “protist”. Generally speaking, protists are eukaryotic microorganisms. In fact, this term does not have a real evolutionary meaning since it includes all eukaryotes that are not animals, plants or fungi. The first registered observation of a protist was done by Leeuwenhoek in 1674, but the term was coined and popularized in 1866 by Haeckel (famous for his detailed illustrations of these organisms, Figure 1), and at the beginning this term also comprised the prokaryotes. *Protista* was considered as a new life kingdom, together with animals and plants, apparently less important and with few categories. Today, thanks to several studies that began in the second half of the XX century, we know that animals and plants are two little leafs in the phylogenetic tree of eukaryotes, which is mostly formed by unicellular forms of life (Figure 2).

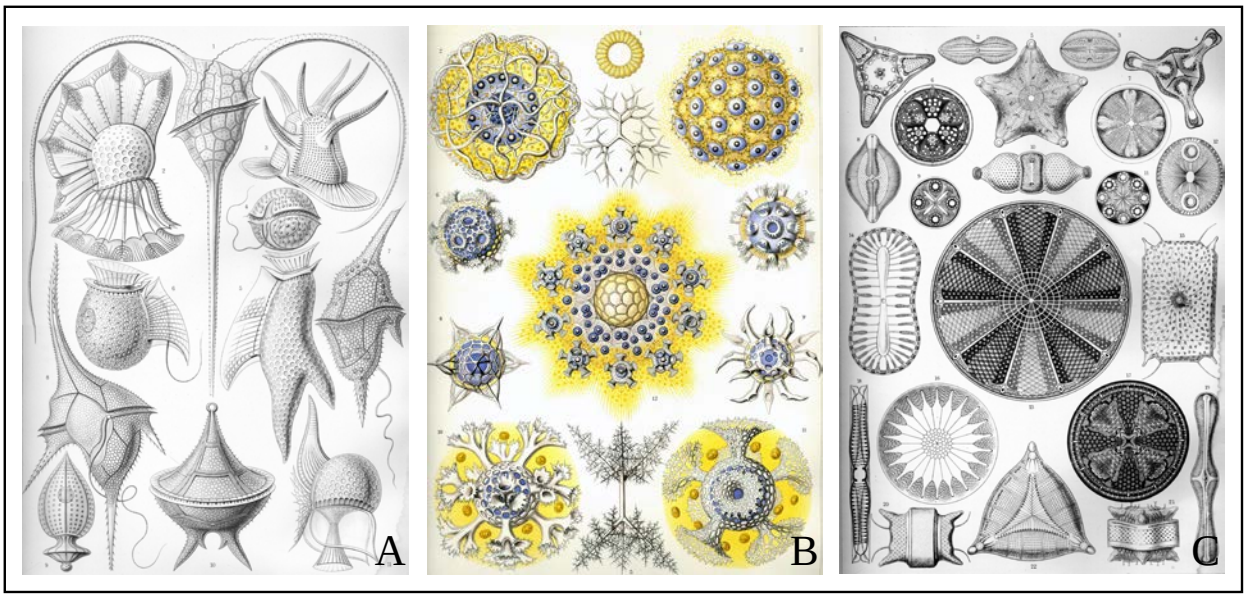

Figure 1. Morphological diversity in Alveolata (A), Rhizaria (B) and Stramenopiles (C). Drawings by Ernst Haeckel, 1904.

Methodological improvements made possible to skip from the study of the visible to the discovery of the invisible. Thus, several techniques are fundamental for the study of microorganisms. A classical method is the observation and counting of protist cells with epifluorescence microscopy, which implies the utilization of cellular stains, as for example DAPI (4’,6-diamidino-2-phenylindole) that binds to the DNA of the cells (Porter and Feig 1980). A more focused technique is Fluorescent In Situ Hybridization (FISH, Pernthaler *et al.* 2003, Massana *et al.* 2006) that targets specific cells using taxon-specific oligonucleotide probes, allows collecting information about the abundance and the global diversity of marine protists (Morgan-Smith *et al.* 2011 and 2013), and can also be used in grazing experiments (Fu *et al.* 2003, Jezbera *et al.* 2005, Massana *et al.* 2009). Another method to quantify microbial abundance is flow-cytometry (Zubkov *et al.* 2006

and 2007; Christaki *et al.* 2011). This technique is really useful with a large amount of samples, but still needs to be improved and optimized for some functional groups like heterotrophic micro-eukaryotes. Despite the possibility to identify some cell types, epifluorescence microscopy is not enough accurate for diversity studies. Indeed, morphology is useless to identify picosized cells (and many nanosized cells), and often does not allow to go deep inside the taxonomy at species level.

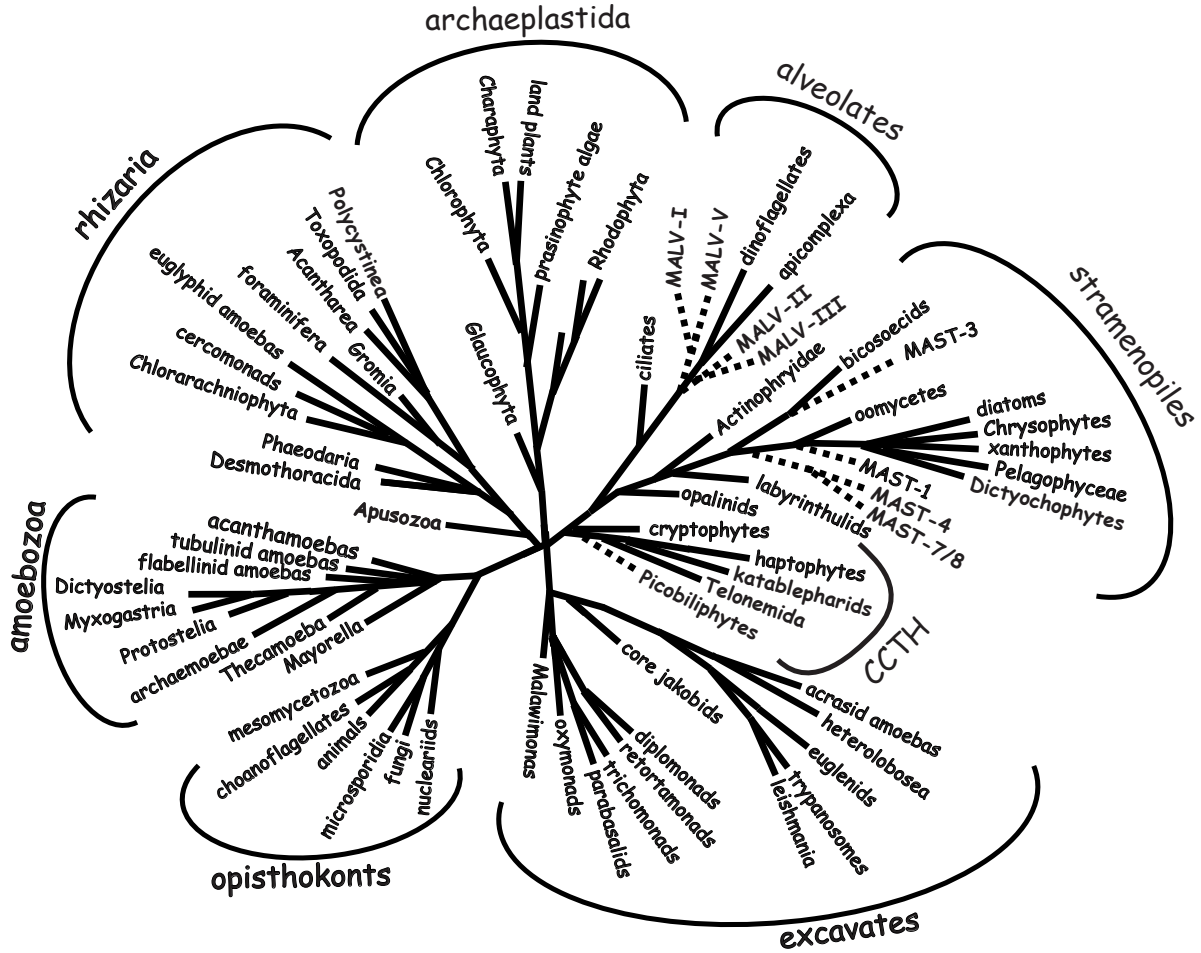

Figure 2. Eukaryotic tree of life, showing the consensus phylogeny of the major eukaryotic groups based on molecular and ultrastructural data (adapted from Baldauf 2003). Dotted lines indicate positions of major lineages known primarly from culture-independent molecular surveys.

With the development of molecular methods the study of microbial diversity improved exponentially. Pioneer protist studies in this direction were those of Díez *et al.* 2001, López-García *et al.* 2001 and Moon-van der Staay *et al.* 2001 targeting picoeukaryotes. These investigations had to face a general problem: to estimate the diversity is fundamental to identify groups of similar organisms that are named "species" in the classical taxonomy. Many different species concepts have been applied to microorganisms in general and protists in particular (Roselló-Mora and Amann 2001, Schlegel and Meisterefeld 2003). The most pragmatic concept proposes that a species is a

"group of organisms that share similar morphological characteristics". The biological concept, perhaps the most useful in animals and plants, defines a species as "a group of organisms capable of interbreeding and producing fertile offspring". Although most protist cell divisions are asexual, sexual reproduction is also known to be present in protists (Amato *et al.* 2007), but there is little information about how spread it is throughout the different protistan groups and how frequent it occurs. Of course, life-cycle studies of protist species are necessary to find out the incidence of asexual-sexual divisions. At present the principal limitation of these life-cycle studies is that only a few protist species are cultured and well-characterized, and even some groups completely lack a cultured representative. So, it is not practical to invoke the biological species concept for studying protist diversity.

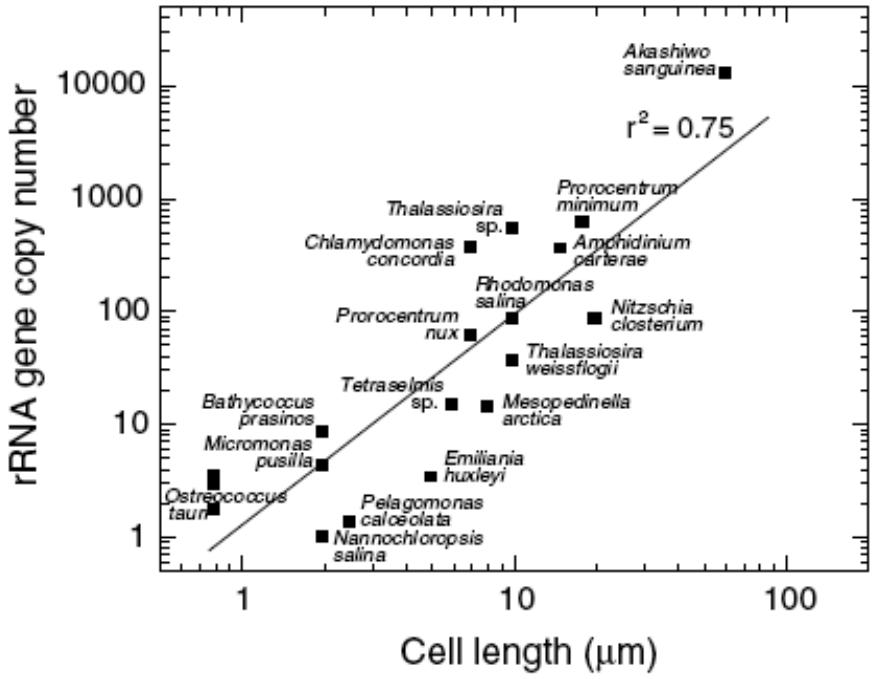

Figure 3. Correlation between cell size and rDNA copy number in different protist species. Taken from Zhu *et al.* (2005).

Luckily for microbiologists, during the 70s of the past century Carl Woese came out with the idea that it was possible to identify and organize all life forms by comparing their DNA sequences (Woese and Fox 1977). This operation needed basically two steps: alignment of DNA sequences of the same gene and measurement of their genetic distances. The preferred target gene for this approach since the beginning was the ribosomal RNA gene (rDNA). This gene is present in all organisms and it is conservative enough to be used in phylogeny among any life form. Molecular taxonomy has several advantages: it can be applied to a wide range of taxa, to all life stages and to

Table 1. Intragenomic 18S rDNA gene variability (SSU).

Reference	Species	Sim	Origin
Rooney *et al.* (2004)	*Cryptosporidium parvum*	92.1	Genome
	Plasmodium Falciparum	89.5	Genome
	Plasmodium berghei	92.0	GeneBank
Alverson *et al.* (2005)	*Skeletonema grethae*	99.2	Strain
	Skeletonema japonicum	99.4	Strain
	Skeletonema menzeleii	99.4	Strain
	Skeletonema pseudocostatum	99.5	Strain
	Skeletonema subsalsum	99.5	Strain
Simon *et al.* (2008)	*Phoma exigua*	99.5	Strain
	Mycospharella punctiformis	99.6	Strain
	Teratospheria microspora	99.6	Strain
	Davidiella tassiana	99.4	Strain
	Aspergillus nidulans	99.6	Strain
Gong *et al.* (2013)	*Tintinnopsis sp.*	99.1	Individual
	Pseudotontonia sp.	99.3	Individual
	Strombidium sp.	99.7	Individual
	Vorticella sp.	99.1	Individual

Values of intragenomic genetic similarity in different microeukaryotes species. Low values of similarity in *Plasmodium spp*. and *Cryptosporidium parvum* are explained by the effective presence of different ribosomal forms activated in different hosts of these parasites.

the large number of data that are typical of most ecological studies (Caron *et al.* 2009). The rDNA is very useful but it is not a perfect target, since it is typically a multi-copy gene, particularly in eukaryotes. In algal strains, the copy number ranges from 1 to 10,000 (Zhu *et al.* 2005) implying that relative gene abundance can deviate strongly from relative cell abundance. The copy number is proportional to cell-size and genome size (Figure 3) so the chances of great variations is lower for pico and nano sized cells. Moreover, it is possible that these copies have some variability at intragenomic level. The risk of intragenomic variability is that we could detect two or more different sequences when there is only one organism. Again, in most cases this intragenomic variability is very low (Table 1).

The important innovation of the molecular techniques was the possibility of a more realistic study of marine microbial diversity, particularly concerning nano- and picosized plankton. The seminal approach was the construction of clone libraries of 18S rDNA genes, which were amplified from environmental genomic DNA by a polymerase chain reaction (Saiki *et al.* 1985) step. Typically, between 100 and 500 sequences were obtained per clone library. These ribosomal sequences became the basis for a new molecular taxonomy; in fact a new “species concept” more pragmatic than the biological or morphological criteria appears: that related to operational taxonomic units (OTUs). Following this criterion, sequences are grouped in countable units that have a certain degree of genetic divergence, chosen by the researcher, in an operation commonly known as clustering. Is important to highlight that the way that sequences are clustered in OTUs is a crucial step

that determines our vision of the diversity in marine samples.

Despite some limitations, the rDNA gene (and particularly the 18S rDNA) is still the best compromise to study protist diversity and has been chosen as target in the emergent high-throughput sequencing (HTS) technology (454 and Illumina), which has evolved so fast that the initial definition of "Next-generation sequencing (NGS)" has become obsolete in less than five years. The number of sequences collected in HTS is several orders of magnitude higher than the one obtained in clone libraries, and then a new problem appears related with the management of these huge amounts of data. However there is a great enthusiasm about the possibility of "sequencing the ocean" (Venter *et al.* 2004) and many scientists are working in the optimization of the methods and improve the confidence of the approach (Kunin *et al.* 2009; Quince *et al.* 2009). High-throughput sequencing gives the possibility to go deep inside in diversity studies. It is important to remember that, when combined with PCR amplicons, HTS is subjected to the same PCR biases (Wintzingerode *et al.* 1997). Nowadays metagenomics, despite having as a principal goal the study of metabolic functions more than species' diversity, is a viable alternative for the collection of 18S rDNA sequences from natural microbial assemblages (Logares *et al.* 2013). The use of metagenomic techniques is independent of the PCR step, so eliminates this source of errors. To better understand the inner characteristics of protist diversity, all the different approaches described have been used in this thesis.

General taxonomy of microeukaryotes

The high-rank taxonomy of eukaryotes at the present time is a continuous matter of debate in the scientific world (Burki *et al*, 2008). Thanks to the combination of microscopy and molecular biology it is possible to identify most taxa, but the real challenge is to understand how these taxa are related among them. In this thesis there is a mix of classical morphological taxa (better defined thanks to molecular tools) and new ribogroups, which are formed by sequences that cluster together in a tree and branch outside the well known groups. So, the new ribogroups are inserted among the classical taxa well defined by morphology, which are used as the backbone of the eukaryotic tree of life. The general taxonomy reference for most morphological groups follows the classification of Adl *et al.* (2012).

Among the entire eukaryotic tree of life (Figure 2) four protists supergroups deserve to be mentioned due to their importance in molecular surveys of protists in the marine environment: Alveolata, Rhizaria, Stramenopiles and CCTH. Alveolata, often the most abundant supergroup, comprises two of the most studied classical classes: dinoflagellate and ciliates. Rhizarian are composed by Radiolaria, which can have solitary or colonial lifestyles and are characterized by complex structures, Cercozoa (also known as Filosa) and Foraminifera, which prefer living in sediments. Stramenopiles encompass phototrophic groups like diatoms, as well as heterotrophic groups such as bicosoecids. The CCTH is a recently proposed supergroup that includes Cryptophyta, Centroheliozoa, Telonemia, and Haptophyta (Burki *et al.* 2009), but more recent phylogenies raises some doubts about its monophyly (Hampl *et al.* 2009, Baurain *et al.* 2010, Burki *et al.* 2012). In the last decade, and thanks to molecular surveys, a "forgotten" group has risen in importance in the oceanic ecosystem, the Fungi (Bass *et al* 2007, Richards *et al.* 2012). For traditional reasons, Fungi were generally studied by botanists and not protistologists. However, since many fungal species are unicellular, they perfectly fit within the microeukaryotes targeted in marine studies. Fungi have been previously found in seawaters, including the deep ocean ecosystem (Bass *et al.* 2007, Jebaraj *et al.* 2009, Edgcomb *et al.* 2011, Richards *et al.* 2012). Initially, fungal sequences were disregarded in protist surveys, like metazoan sequences, but now they are appreciated and kept. In fact, it does not make sense to exclude such important marine players.

Recently defined ribogroups represent a great part of retrieved sequences in marine molecular surveys. In fact the majority of the sequences of this thesis belong to Marine Alveolates (MALV) that were already detected in the first molecular surveys of deep marine waters (López-Garcia *et al.* 2001) and better defined later (Groissillier *et al.* 2006). Other important ribogroups are the Marine Stramenopiles (MAST), defined in 2004 by Massana *et al.* and the Picozoa (known before

as Picobiliphyta) that were first identified by environmental sequences (Not *et al.* 2007a) and later cultivated (Seenivasan *et al.* 2013). In Rhizaria there are three ribogroups, RAD-A, RAD-B, and RAD-C, the second encompassing the former morphological group of Sticholonche previously known as Taxopodia. All these ribogroups are now widely found and recognized, therefore entering de facto in "practical" taxonomical schemes.

Trophic roles and participation in biogeochemical cycles

In one millilitre of epipelagic seawater there are about 1000-10,000 cells of microeukaryotes. It is difficult to identify the ecological function of each different taxa but it is clear that together they play important roles in biogeochemical cycles, both as autotrophs and heterotrophs. It is worth to remember that phytoplankton, today considered as the sum of phototrophic bacteria and phototrophic protists, produces 70% of the total oxygen of the planet (Epstein *et al.* 1993) and makes life on Earth possible. Generally unicellular organisms are connected in a complex size-based trophic webs. The microbial loop, proposed by Azam *et al.* in 1983, constitutes an interesting hint of this web. Through this loop, the dissolved organic matter is consumed by prokaryotes and arrives to upper trophic levels thanks to the fact that phagotrophic protists feed on prokaryotes and are then fed by larger zooplankters (Figure 4 and Figure 5b). Prokaryotes could be considered as the biochemical machines that drive the principal biogeochemical cycles (carbon, nitrogen, sulfur), but at the end who controls the velocity of these metabolic reactions are the bacterivorous protists, probably together with viruses (Boras *et al.* 2010).

Phagotrophy, the ingestion of food particles through engulfment of the cell membrane, is widespread among protist taxa. Both in cultured species and in environmental samples (mostly in epipelagic waters) is a quite well studied process. Important grazer classes are for example the ciliates, which are the predators in the classical food web, and the chrysophytes or the MAST-4 ribogroup, perhaps the most important bacterivorous in the marine microbial loop (Massana 2011). Nevertheless, phagotrophy is not the only form of heterotrophy in the environment. Several species belonging to Fungi (Richards *et al.* 2012), Excavata (von Der Heyden *et al.* 2004), Chrysophyceae (Holen and Boraas 1996, Sanders *et al.* 2001) or Labyrinthulidae (Raghukamar *et al.* 2001) survive by osmotrophy, which is the uptake of dissolved organic compound by osmosis. However the prevalence of this phenomenon is still not well understood. In addition, there are several examples of groups that survive in the ocean thanks to parasitic interactions with a varied array of marine hosts. These marine parasites include the MALV-I and -II ribotypes, the most abundant groups in terms of sequences retrieved (Siano *et al.* 2010), pirsonids (Schnepf *et al.*

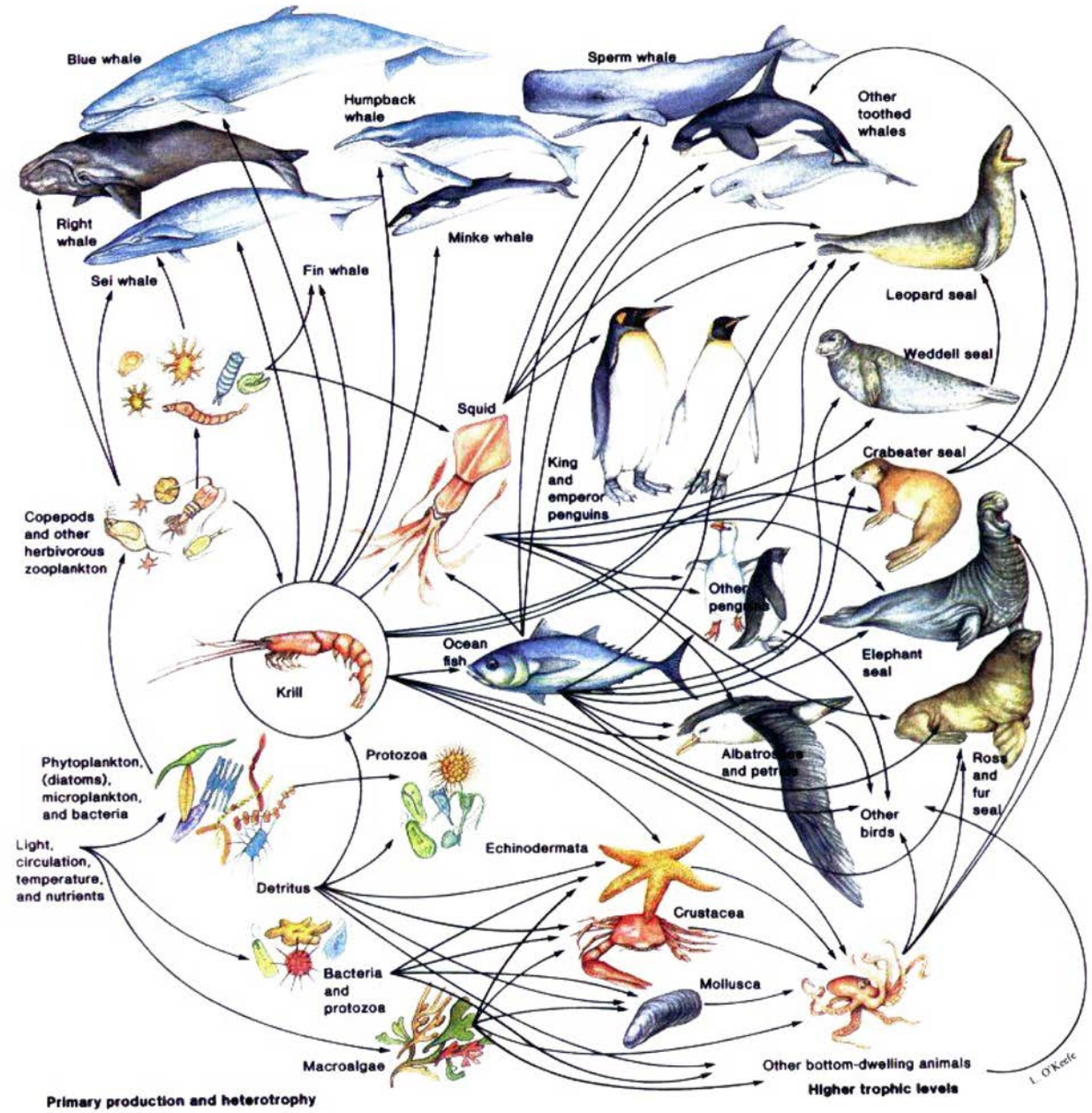

Figure 4. A complete marine food web, indicating a large array of species and their interactions. The microbial food-web is shown on the left, and highligths the trophic connections between microbiota and macrobiota. Drawing by O'Keefe L.

1990) and several fungal species (Richards *et al.* 2012). The environmental study of parasitism is quite hard and despite a few documented cases (Chambouvet *et al.* 2008), the information about the magnitude of the global phenomenon is, as for osmotrophy, poorly understood. Apart from these strict trophic divisions, it is important to remember that the unicellular world seems to favor a sort of plasticity in the trophic style, and mixotrophy, the combination of autotrophy and heterotrophy in the same organism, appears as a common behaviour between protists taxa (Sanders *et al.* 1991, Jones 2000, Zubkov *et al.* 2008).

The deep ocean: a peculiar habitat

The ability of adaptation of microeukaryotes is evident by the fact that they are widespread in the planet, including all sorts of extreme environments. In the ocean, we know that they are present in the entire water column (Not *et al.* 2007b). However, for obvious reasons, surface communities are much better studied than deeper ones. In fact, the functioning of the deep ocean ecosystem is far from being completely clear. Traditionally the deep dark ocean is divided in three zones, the mesopelagic (200-1000 m), the bathypelagic (1000-4000 m) and the abyssopelagic (more than 4000 m). The mesopelagic layer, where often resides the thermocline, appears to be more influenced than the other two deeper layers by the epipelagic system (0-200 m). Indeed, a large fraction of the organic carbon fixed by photosynthesis is respired in this zone (Aristegui *et al.* 2005).

The bathypelagic zone shows several differences compared with upper ecosystems. Considering physical parameters, this system is more stable: water are generally well oxygenated (although anoxic basins exist), temperature exhibits a very narrow range globally, from 1 to 4 ºC, and salinity is practically constant at 35 ppm. In the bathypelagic region the pressure is really high (5 to 10 MPa), but this is not limiting the development of life at macro and micro scale. Despite this apparent homogeneity, it is still possible to recognize several different water masses based on the physical and chemical parameters, being the most important the North Atlantic Deep Water (NADW), the Circumpolar Deep Water (CDW) and the Weddel Sea Deep Water (WSDW). These are found principally in the Atlantic, Pacific and Indian Oceans, respectively.

From a chemical point of view, the concentration of organic matter, inorganic nutrients and other chemical compounds can be very different in different marine regions, depending on the sinking material from the surface. Generally, the bathypelagic ocean is rich in the oxidized forms of inorganic nutrients (NO_3, PO_4) and is depleted of reduced compounds such as ammonium (Nagata *et al.* 2010). Globally the deep ocean is considered the largest reservoir of bioavailable organic carbon (Libes 1992; Benner 2002), and the concentration of dissolved organic carbon (DOC) differs among the different basins (Hansell and Carlson 1998). The role of deep ocean as inorganic carbon sink is quite intuitive (Figure 5a). Between 5 and 15% of the carbon fixed by photosynthesis in the upper marine layer sinks to the bathypelagic realm through the biological pump (Giering *et al.* 2014), where is respired and sequestered for centuries until returned to the upper ocean and then to the atmosphere. Thus, the bathypelagic system has an extremely important role in the global balance of CO_2 and, considering the link of this balance with critical problems such as global warming and climate change, is really important to define the final destination of deep DOC (Aristegui *et al.* 2009).

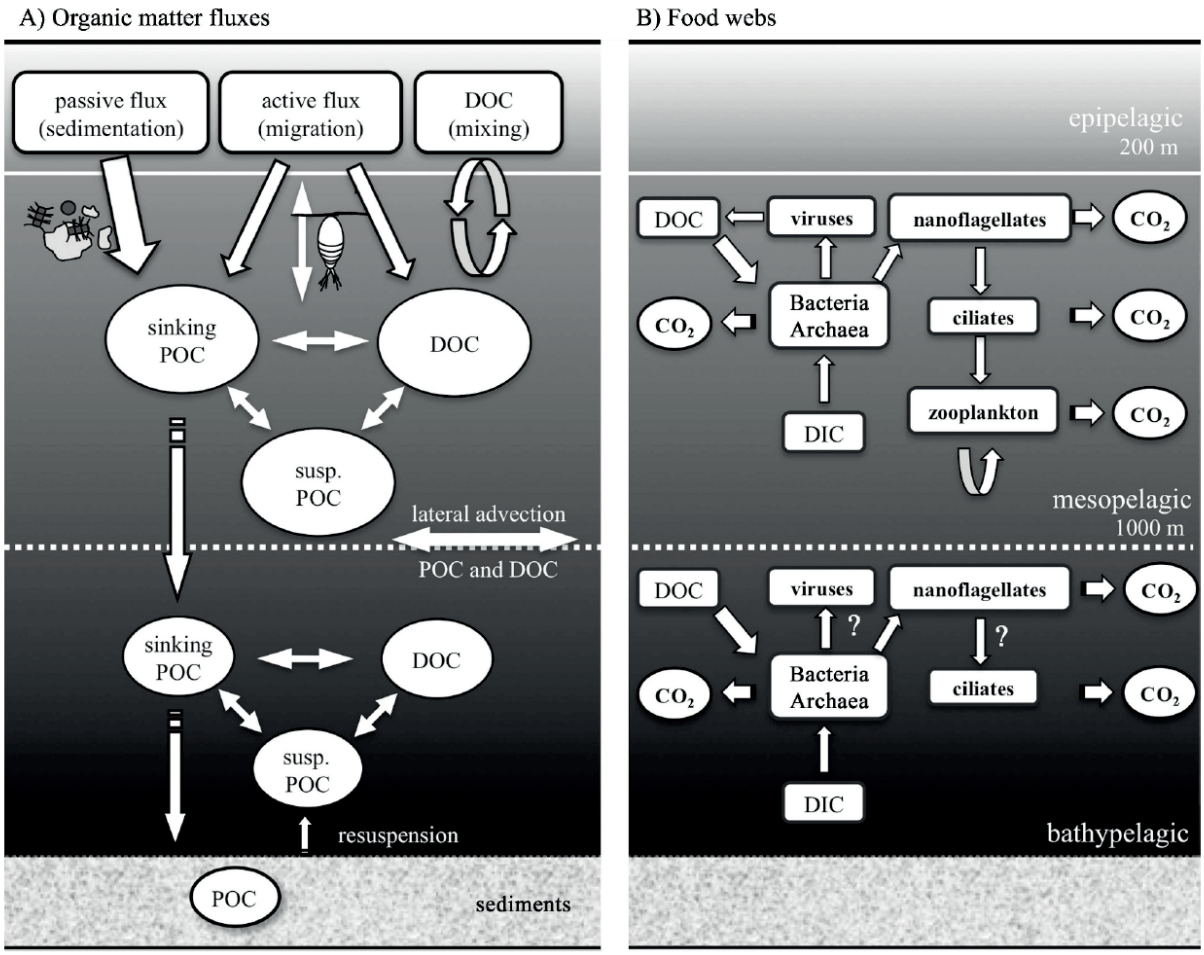

Figure 5. Schematic representation of organic matter fluxes (A) and the microbial food web (B) in the oceanic deep ecosystem (from Aristegui *et al.* 2009). A) Three interconnected pools of organic carbon are indicated: dissolved organic carbon (DOC), sinking particulate organic carbon (POC) and suspended POC. B) Microbial trophic web of the mesopelagic and bathypelagic realms. Prokaryotes in the dark ocean may take up DOC (heterotrophy) and inorganic carbon (chemosynthesis). In the bathypelagic zone prokaryotic control by flagellates or viruses and the role of ciliates remain enigmatic (question marks).

At the bathypelagic level there is a general decrease of DOC concentration along the path of the deep global thermohaline circulation (Figure 6). The concentration of DOC is high (>50 µmol kg^{-1} C) in the newly formed North Atlantic Deep Water (north of 50ºN), tends to be a bit lower and constant in equatorial regions (about 45 µmol kg^{-1} C), and further decreases in the south to a minimum of 39 µmol kg^{-1} C. The constant DOC concentration along South Indian Ocean (40 µmol kg^{-1} C) suggests a net carbon input, due to the invasion of circumpolar deep water (CDW), and then a subsequent removal. Bottom waters of the Pacific Ocean gradually lose organic carbon as they move northward: DOC is 42 µmol kg^{-1} C in the circumpolar waters of south Pacific and decreases to 36 µmol kg^{-1} C as the water slowly enters the deep North Pacific (Figure 7, Hansell *et al.* 2009). Probably this decrease of the DOC is explained by biological consumption by heterotrophic prokaryotes and perhaps fungi.

The abiotic characteristics of the deep ocean defines a habitat really different from the surface one.

There are fragmentary information about the abundance and distribution of microeukaryotes in the dark ocean water column (Tanaka and Rassoulzadegan 2002, Yamaguchi *et al.* 2004, Fukuda *et al.* 2007, Sohrin *et al.* 2010, Morgan-Smith *et al.* 2011, Morgan-Smith *et al.* 2013). Diversity studies have been performed in the water column (López-Garcia *et al.* 2001, Stoeck *et al.* 2003, Countway *et al.* 2007, Not *et al.* 2007b), in sediments (Edgcomb *et al.* 2011, Salani *et al.* 2012) and in marine chimneys of hydrothermal vents (Edgcomb *et al.* 2002, Sauvadet *et al.* 2010). From these papers we know that the diversity of deep water protists appears dominated by Alveolata and

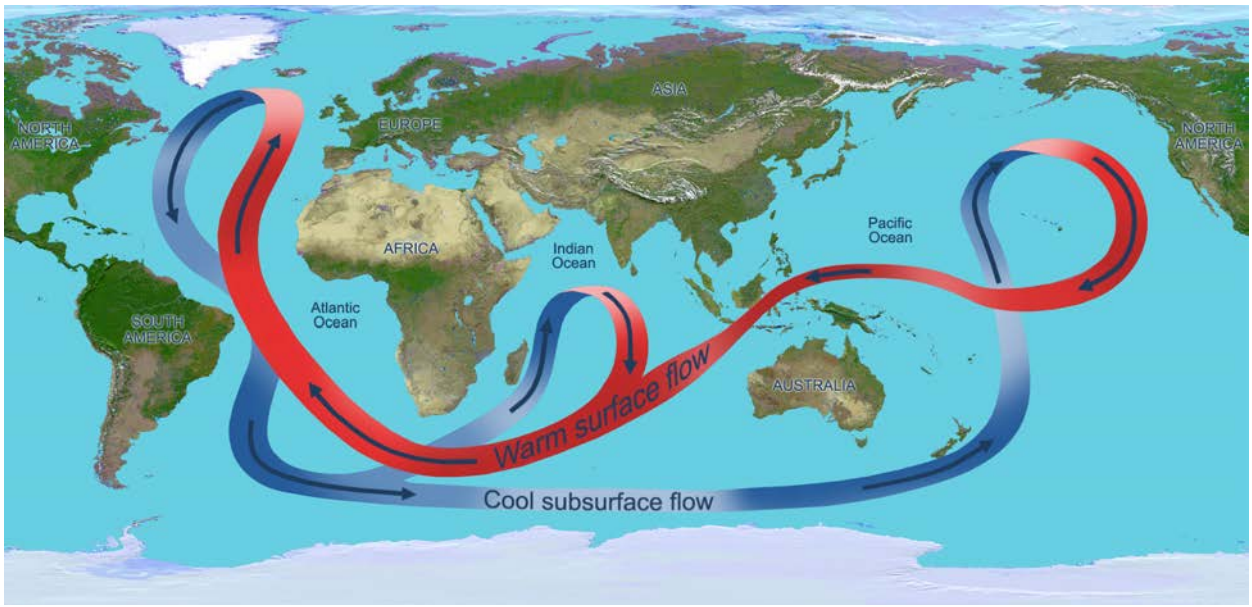

Figure 6. Thermohaline circulation. Cold and dense water masses sink in the North Atlantic and Southern oceans, creating a current that flows in the ocean basins. These waters return to the surface thanks to upwelling events in Indian and North Pacific oceans, forming a current of warm water that flows in the opposite direction in upper layers. Near the North Pole the water get colder and sink restarting a cycle that lasts centuries.

Radiolaria whereas Fungi dominate in sediments. Despite not being one of the dominant groups, Excavata apparently prefer deep waters than surface. As phototrophy is not possible in the dark ocean, the community of bathypelagic protists should present one the three previously mentioned heterotrophic lifestyles: phagotrophy, osmotrophy or parasitism. The relative importance of each trophic mode at the ecosystem perspective is still a matter of debate.

To achieve a global vision of the functioning of the deep ocean, the Malaspina circumnavigation cruise was performed in 2010 on board the R/V BIO_Hesperides. This cruise started in Cadiz (Spain) and sampled 147 stations all over the world (Figure 8). The principal aim of this expedition was the study of the dark ocean at a global scale, including data about microeukaryotes. The magnitude and multidisciplinarity of the sampling effort allowed us to compare our data with parallel parameters in order to achieve a more complete understading of the entire system.

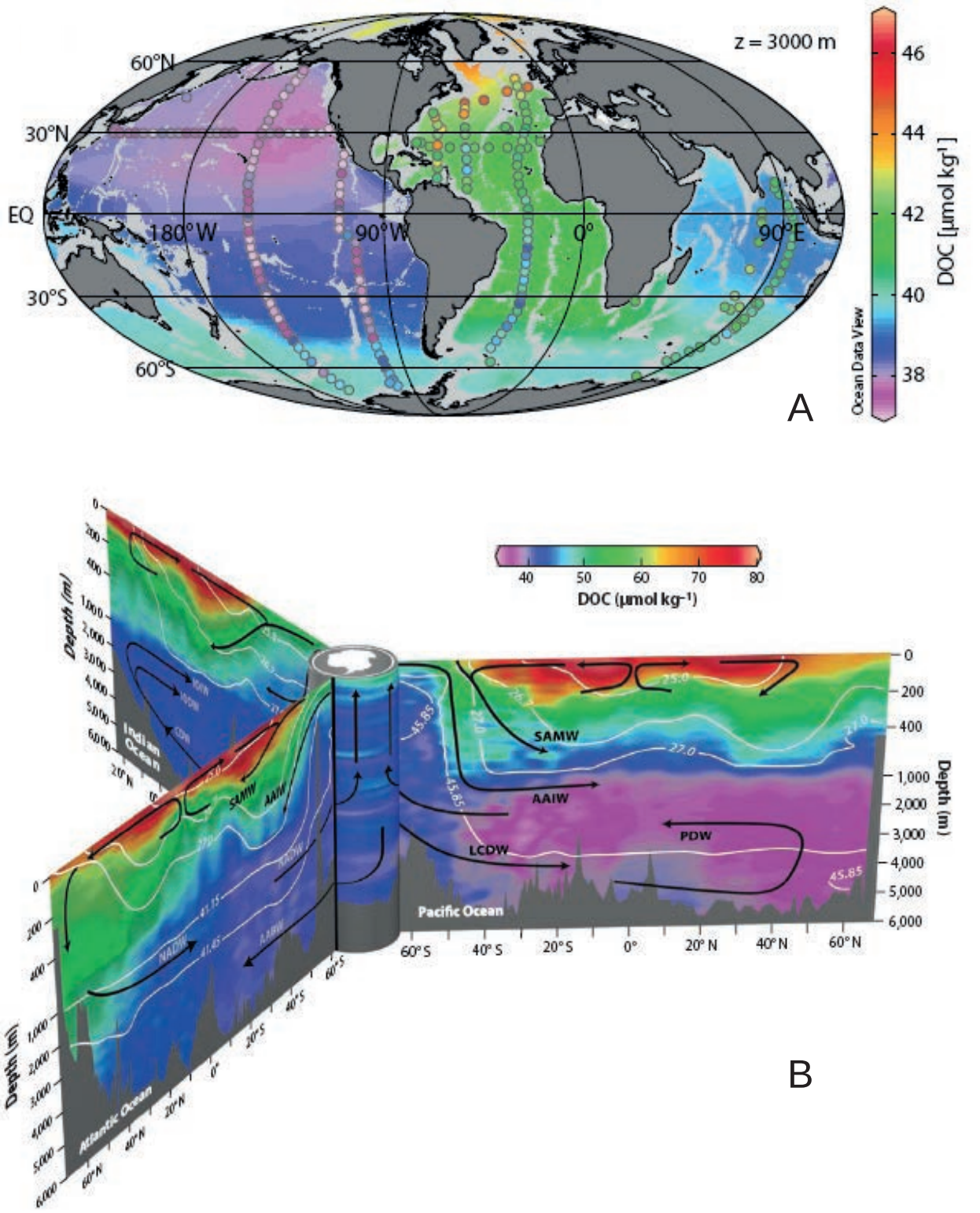

Figure 7. Distribution of dissolved organic carbon (DOC; μmol Kg^{-1}) in the global ocean (Hansell *et al.* 2009). A) Distribution of DOC at 3000 m. Dots are observed values, while the background field is modelled. B) Distribution of DOC in the central Atlantic, central Pacific and eastern Indian ocean. Arrows depict water mass circulation.

Aim of the thesis

The general aim of this thesis is to draw a global picture of the community of marine microeukaryotes. The achievement of this goal was structured in four chapters. The first chapter (*Sequence diversity and novelty of natural assemblages of picoeukaryotes from the Indian Ocean*, ISME J. 2011), the study of the diversity of epipelagic community through clone libraries, was useful as a first approach to molecular biology tools and to establish a guideline on how to treat sequence datasets (i.e. alignment, clustering threshold, diversity estimates). In the second chapter (*General patterns of diversity in major marine microeukaryote lineages*, PLOSONE 2013) sequences derived from all the reports published before 2010 were analyzed in order to describe several features of the genetic diversity of microeukaryotes groups. Moreover, an explorative study of

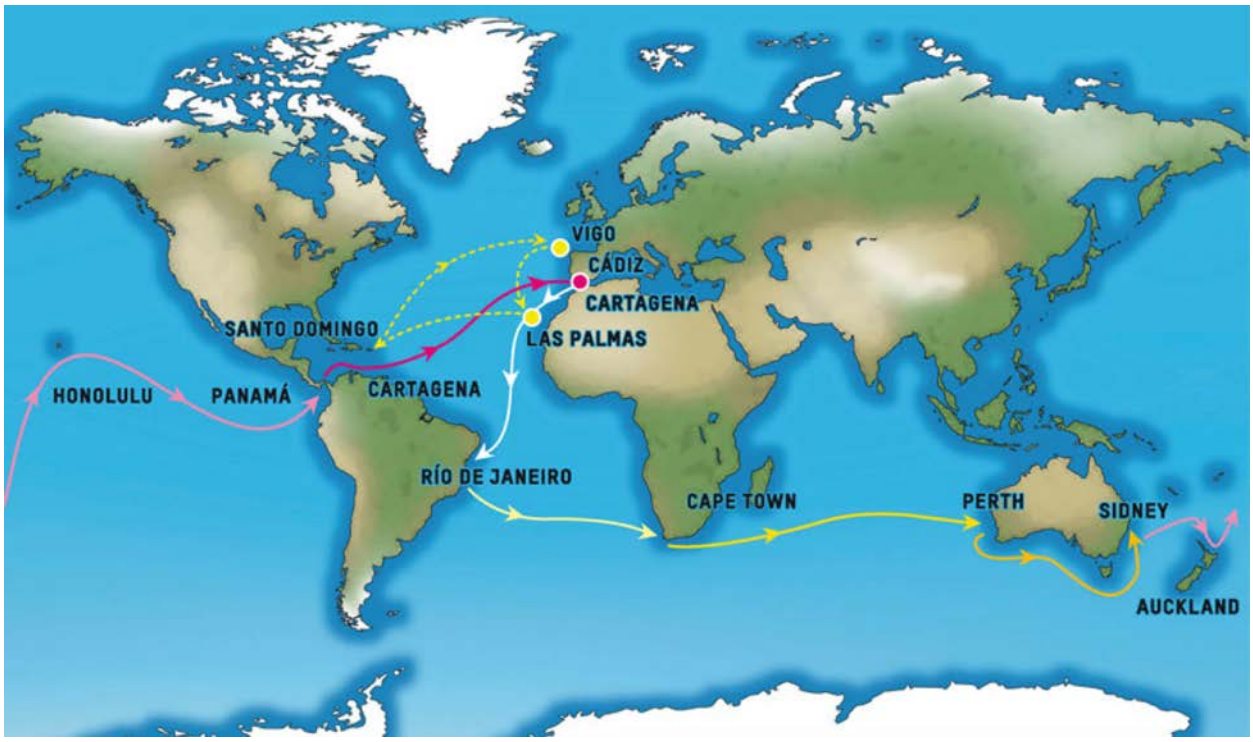

Figure 8. Cruise itinerary of the Malaspina 2010 expedition, including the tracks from the ship Hesperidés (continuous line) and Sarmento de Gamboa (dotted line).

the evolutionary model for the different taxa was performed. The most precious fruit of this work was a well annotated set of sequences, all belonging to the V4 region of 18S rDNA, which were the core for a reference database (MAS9013) used for taxonomic identification and chimera check in the successive studies done by pyrosequencing the same rDNA region. The second part of the thesis, in the frame of Malaspina-2010 project, was focused on the deep ocean ecosystem. The third chapter (*Global abundance of planktonic heterotrophic protists in the deep ocean*, submitted to ISME J.) investigates the abundance of heterotrophic flagellates in the global meso- and bathy-

pelagic regions with the combined use of epifluorescence microscopy and flow-cytometry. In the fourth chapter (*Diversity of marine microeukaryotes in the global deep ocean*, in preparation), we studied the phylogenetic diversity and biogeography of microeukaryotes, and their relation with environmental parameters at the boundary between bathypelagic and abyssal regions through rDNA pyrosequencing and metagenomic approaches.

The outline of different topics studied can be explained under two general objectives and several specific ones, as follows:

Objective 1: *Defining the taxonomic groups of marine microeukaryotes and their genetic structure*

The first part of the thesis represents an effort to define our "diversity unit" from studies based on the well-known molecular cloning and Sanger sequencing in order to establish a solid base for the second part of the thesis. We started with data from one cruise (Chapter 1) and then continued with the analysis of the complete 18S rDNA database available at that time (Chapter 2). The specific objectives of this part were:

- To select the region of the 18S rDNA gene that best represents the variability of the complete gene

- To identify a reasonable similarity threshold for OTU clustering

- To establish the maximal distance in groups delimited at a class-rank level

- To highlight the typical taxonomic classes forming surface communities

Objective 2: *A descriptive study of global deep ocean communities*

The Malaspina expedition allowed us to have a comprehensive set of samples coming from all over the world with associated abiotic and biotic parameters. Such a large amount of data was the base for studying deep microeukaryotes (Chapters 3 and 4) following the next specific objectives:

- To determine the abundance, biomass and distribution of microeukaryotes in the water column between 200 and 4000 m depth

- To study the diversity of bathypelagic microeukaryotes through pyrosequencing and metagenomics approaches

- To identify the abiotic and biotic parameters explaining the abundance and diversity of deep microeukaryotes, with a particular emphasis on the relation with prokaryotes